

Tagmentation on Microbeads: Restore Long-Range DNA Sequence Information Using Next Generation Sequencing with Library Prepared by Surface-Immobilized Transposomes

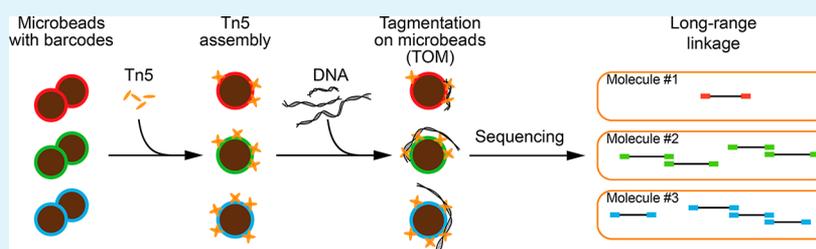
He Chen,[†] Jiacheng Yao,[‡] Yusi Fu,[†] Yuhong Pang,[†] Jianbin Wang,^{*,‡,§} and Yanyi Huang^{*,†,§} 

[†]Beijing Advanced Innovation Center for Genomics (ICG), Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[‡]School of Life Sciences, and Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China

[§]College of Engineering, and College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Supporting Information



ABSTRACT: The next generation sequencing (NGS) technologies have been rapidly evolved and applied to various research fields, but they often suffer from losing long-range information due to short library size and read length. Here, we develop a simple, cost-efficient, and versatile NGS library preparation method, called tagmentation on microbeads (TOM). This method is capable of recovering long-range information through tagmentation mediated by microbead-immobilized transposomes. Using transposomes with DNA barcodes to identically label adjacent sequences during tagmentation, we can restore inter-read connection of each fragment from original DNA molecule by fragment-barcode linkage after sequencing. In our proof-of-principle experiment, more than 4.5% of the reads are linked with their adjacent reads, and the longest linkage is over 1112 bp. We demonstrate TOM with eight barcodes, but the number of barcodes can be scaled up by an ultrahigh complexity construction. We also show this method has low amplification bias and effectively fits the applications to identify copy number variations.

KEYWORDS: *microbeads, surface, tagmentation, sequencing, long-range information*

INTRODUCTION

Next generation sequencing (NGS) technologies have been rapidly developed in the past decade and have revolutionized current research and applications in life sciences and medicine.^{1,2} For most NGS methods, especially those prevalent in high-throughput approaches, it is challenging to provide long-range DNA information because of the limitation in library size and read length. Such characteristics may hinder various applications in which the long-distance relationship of the DNA sequences is important—repeat-region sequencing,³ genome assembly,⁴ haplotype phasing,⁵ and alternative splicing analysis to name a few.⁶ Although single-molecule sequencing technologies are promising on long read-length,^{7,8} their sequencing accuracy and throughput are still insufficient.^{9,10} A practical approach to filling such genomic information gaps is to encode additional information in each short NGS read to retain the long-range relationship between reads. Matured and widely used examples include paired-end,¹¹ mate-pair,¹² HiC,^{13,14} synthetic long read,^{15,16} and several microfluidic-based technologies.^{10,17–19} In general, however, these experi-

ments are labor-intensive and cost-inefficient. Therefore, there is a strong need for a high throughput, cost-effective, and versatile method to recapture such inter-read information in NGS.

Recently, hyperactive Tn5 transposase-mediated library preparation methods have shown great potential in simplifying the experimental process by performing both DNA fragmentation (cleaving) and adapter ligation (tagging) simultaneously.^{20,21} Two Tn5 transposases with each binds to a double-strand oligonucleotide containing the 19-bp mosaic end (ME) could be in vitro assembled into an active dimeric Tn5 transposome.²² The constructed Tn5 transposome can randomly fragmentize template DNA and ligate the oligonucleotides onto the ends of fragmented DNA.²³ This tagmentation process is proven to be highly efficient and rapid, with low starting material requirement.²⁰ Various

Received: January 27, 2018

Accepted: March 15, 2018

Published: March 15, 2018

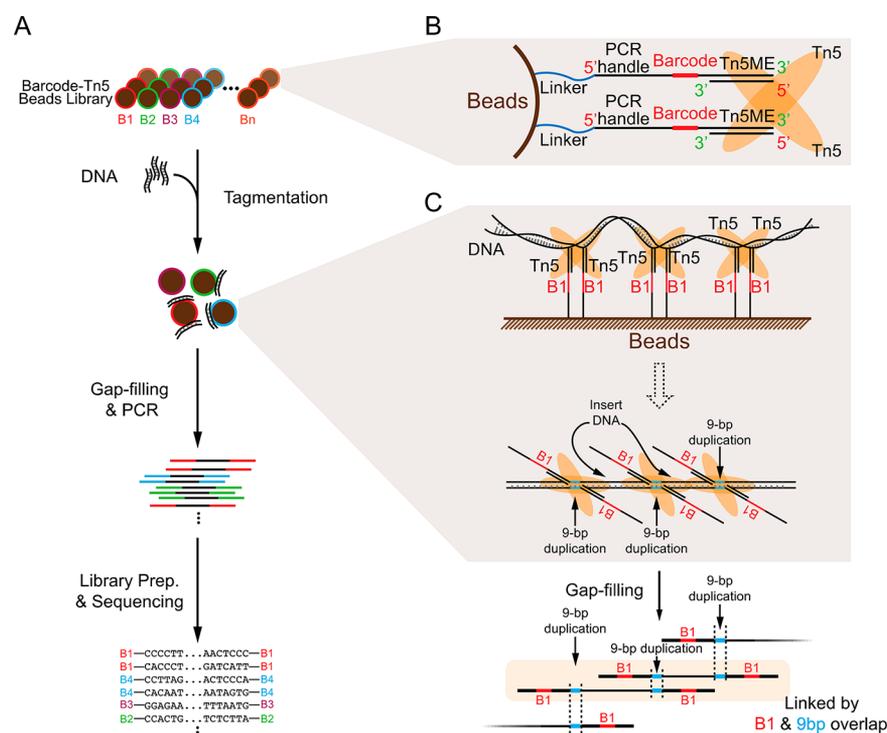


Figure 1. Overview of the tagmentation on microbeads (TOM) method. (A) Experimental procedure for library generation by barcoded microbeads tagmentation. (B) The illustration of the Tn5 transposome on barcoded microbeads. (C) Linkage between neighbor fragments. During tagmentation, each template DNA molecule was tagmented by one bead. Hence, fragments that were generated from the same DNA molecule would have identical barcode. Neighbor fragments also share the same 9-bp duplication at the cut-site.

applications have been developed on the basis of Tn5 tagmentation and sequencing. For example, ATAC-seq maps the transposase-accessible chromatin region and provides genome-wide open chromatin profiles even for single cells.²⁴ With the capability to label template DNA with specific oligonucleotides, the tagmentation process is an ideal editing tool to attach barcodes to each fragment thus leaving the information on the template DNA. Amini et al. developed CPT-seq which uses the contiguity preserving Tn5 tagmentation and combinatorial indexing to capture the haplotype information.²⁵ However, this method requires sufficient resources (e.g., abundant Tn5 transposase, custom sequencing recipe and primers) and it is difficult to be adapted to other sequencing applications.

In this work, we report a tagmentation on microbeads (TOM) strategy to restore long-range DNA sequence information using next generation sequencing with library prepared by surface-immobilized transposomes. We use microbeads-attached identical-barcoded oligonucleotides and encode the information to template DNA through tagmentation. With each unique barcode representing the original DNA molecule, the long-range DNA information is retained. We did a proof-of-concept experiment with eight barcodes and verified the existence of this long-range linkage in the sequencing data. TOM successfully restored the linkage information and could be used on both 5'- and 3'-end tethered beads with high efficiency and low bias. It can also be scaled up to construct ultrahigh complexity barcodes for further applications.

RESULTS AND DISCUSSION

In conventional NGS experiment, DNA fragments are sequenced independently. To restore the long-range information, a connection between reads has to be built. We took

advantage of the barcoding ability of Tn5 transposome. During the tagmentation process, Tn5 transposome cuts the template DNA and adds the oligonucleotides with barcodes to both ends of the fragmented DNA, which would be presented in different fragment ends (reads) during sequencing. This feature provides the possibility to link two “neighbor” fragments that cut by one Tn5 transposome, because they share the two barcodes from that single Tn5 transposome. However, barcodes with ultrahigh complexity were required to distinguish every fragment, which in theory should be higher than the number of fragmented molecules ($\sim 10^{10}$ for 5 ng genomic DNA), making this approach hard to implement and extremely expensive. To reduce the barcode complexity requirement, we took advantage of another intrinsic feature of Tn5 transposition, a 9-bp target sequence duplication after tagmentation and repairing.²⁶ Such 9-bp end-duplications are present at both ends of molecules after in vitro gap-filling and can be treated as natural extensions of barcodes. Because the Tn5 cut-sites are randomly distributed along the template DNA,²⁰ it is unlikely that two molecules share the same cut-site by coincidence. Thus, fragments with the same barcode but different 9-bp end-duplications should not have been cut by the same transposome. Theoretically, combining the barcode and the 9-bp end-duplication sequence as a unique fragment identifier (UFI) can accurately identify the neighbor fragments with barcode complexity 5 orders of magnitude ($4^9 = 262\,144$) lower than the theoretical requirement. Ideally, with 100% reaction efficiency and recapture rate of all fragments, we could use UFIs to concatenate every fragment to reconstruct the original full-length template DNA molecules, only requiring the barcode complexity to be larger than the count of the most abundant fragment. The one essential requirement for this approach is that the two barcodes from one transposome should be either the same or with

certain linkage. However, such requirement is difficult to achieve in in-solution reaction but feasible when barcodes are immobilized on solid-phase.

We designed a Tn5 tagmentation strategy with solid-phase immobilized barcoded oligonucleotides to recover long-range information and reduce the requirement for high complexity barcodes. The overall experimental process of our TOM method (Figure 1A) contains two major steps: (1) use solid-phase immobilized barcoded oligonucleotides with Tn5 transposase to tagment the DNA fragments; (2) amplify the DNA fragments with PCR, followed by standard NGS procedures. The first step is critical, requiring careful design of the surface immobilized oligonucleotides (Figure 1B). Each bead is linked, through biotin–streptavidin binding, with $10^5\sim 10^6$ copies of DNA oligonucleotides with identical sequence. The oligonucleotide contains several parts: a 5'-biotin modification, a linker to reduce steric effect, a PCR handler to amplify the fragment with barcodes after tagmentation, a 12-bp barcode, and a 19-bp hyperactive Tn5 mosaic end sequence (Tn5ME) that is necessary for transposome construction.²¹ To ensure each DNA molecule was tagmented by only one bead, which will result in identical barcode representing one molecule, beads with a comparable number of input DNA molecules were added to the reaction ($\sim 10^8$ in each reaction). If more than one DNA molecules were tagmented by one bead, these DNA molecules were unlikely to be from the same genomic region, and they could be distinguished by their genomic position when aligned to the reference genome. Also, considering the input DNA length (~ 20 kb or ~ 6.8 μm), we used beads with comparable circumference (2 μm in diameter). The small diameter of the beads can help keeping beads in suspension to reduce the probability of tagmentation occurring on multiple beads. After tagmentation and gap-filling, the neighbor fragments will have the same barcode and 9-bp end-duplication that provide reliable linkage between fragments (Figure 1C and Figure S1).

As a proof-of-principle test, we started with 5 ng of genomic DNA from a human cell line HEK293 as input to verify whether the barcodes combined with the end-duplication can be used as the UFIs to link the neighbor fragments in a low-complexity-barcode scenario. We first synthesized oligonucleotides with eight different barcodes and immobilized them on beads separately. With $10^5 \sim 10^6$ copies of identical barcoded oligonucleotides on each bead, we merged equal portions of eight beads to form a pool (identical-on-bead). By adding Tn5 transposase to the beads pool, Tn5 transposome were assembled. Using such bead-immobilized Tn5 transposome for tagmentation, each fragment from one template DNA was linked on the same bead with the same barcode through phosphodiester bond.²⁷ The tagmentation reaction was conducted for 60 min to maximize the tagmentation efficiency. Because of the magnetic beads employed, the reaction can be terminated by changing buffer after magnet separation without requirement for purification. Then 12–18 cycles of PCR were performed to enrich the fragments with barcodes at both ends, followed by a standard NGS library preparation with slight modification.

Paired-end sequencing was performed to obtain barcodes at both ends of insert DNA. The barcodes could be easily recognized from the sequencing reads through their characteristic positions and sequence context. We then examined the barcodes of each pair of reads to check whether each template DNA molecule was tagmented by one bead. If both reads in a

pair had identical barcodes, this fragment was either tagmented by a single bead, or tagmented by two beads with the same barcode which was a rare event through increasing the variety of barcodes. In the sequencing data, we observed that majority of the paired-end reads (>85%) had the same barcode, which indicates each template DNA was tagmented by one bead. As a control experiment, we immobilized eight different barcodes on one bead pool (mixed-on-bead). In this way, the barcodes in the paired-end reads were randomly arranged after tagmentation and sequencing. As expected, a small portion of the paired-end reads (<13%) had the same barcode, which was the result from random assignment (Figure 2A). These observations

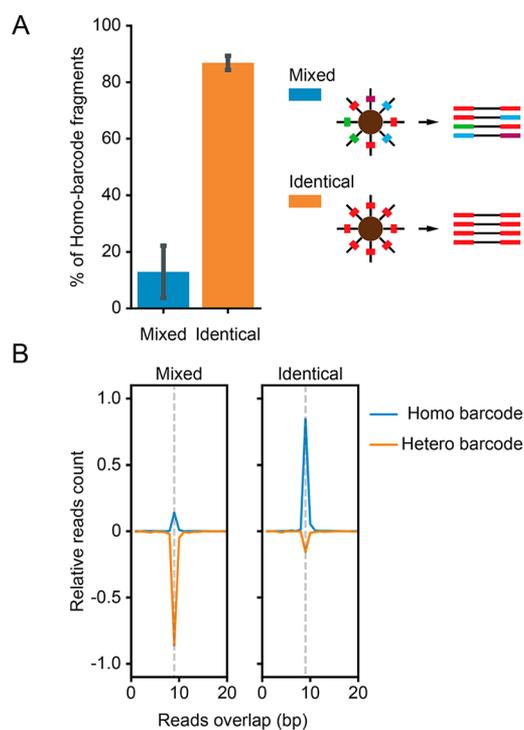


Figure 2. Long-range linkage in proof-of-principle experiment. (A) Percentage of the fragments that have identical barcodes in both ends while using two types of beads. We constructed mixed beads by mixing eight different barcoded oligonucleotides and tethering them on beads. Identical beads were constructed by first tethering identical barcoded oligonucleotides on beads then mixing eight kinds of beads together. (B) Relative counts for overlapped fragments. For each fragment, we search for overlapping reads within its 20 bp neighborhood. We sorted all fragments with such overlapping reads by number of overlapped bases, and we examined if the overlapped ends had the same barcode (Homo) or not (Hetero). The 9-bp peak was clearly indicated by gray dash line. The counts were normalized by the total counts with 9-bp overlaps. The Hetero group counts were multiplied by -1 for better illustration.

proved that tagmentation for each template DNA molecule indeed occurred on a single bead. Therefore, using TOM, we can encode barcodes into DNA molecules as unique molecule identifiers (UMIs) to label and quantify molecules.

The application of UMI is limited if the number of barcodes is smaller than the number of identical molecules. Low barcode complexity is not sufficient to determine the origin of fragments because fragments from different template DNA molecules could share the same barcode. Thus, it still requires high barcode complexity. Recently, Zhang et al. developed a method call CPTv2-seq using the same principle to resolve haplotype

phasing by a beads library with $\sim 150k$ barcodes,²⁸ which is not feasible for most researchers. Conventionally, the genomic position (Tn5 cut-site in our case) can also be considered as a barcode. So, in our low-barcode-complexity scenario, when the barcode is combined with the 9-bp end-duplication at the cut-site for each fragment end, some linkage between the neighbor fragments might be restored. We first confirmed the 9-bp end-duplications exist in the NGS sequencing data by comparing the fragments prepared by tagmentation in solution-phase Tn5 transposome with those by sonication.³ The signature of 9-bp overlap only existed in the transposase treated libraries, and 6.7% of the fragments were overlapped with 9-bp duplications (Figure S2A). We then analyzed the fragments that were prepared by TOM, to confirm these overlapped ends share the same barcode. We found a large majority (88.4%) of these overlapped fragments contained the identical barcode at the overlapped ends, compared to 16% in those control conditions that using mixed barcodes on beads (Figure 2B and Figure S2B). These results proved that through TOM, fragments generated from the same molecule would have identical barcode and also share the same 9-bp cut-site duplication with their neighbor fragment at both ends. This finding offers a unique potential to concatenate the adjacent short DNA fragments into a long one. The fragments from different molecules but sharing the same 9-bp end-duplication by chance could be recognized by examine the barcodes. The small portion of overlapped fragments that had different barcodes at the overlapped ends were generated by different template DNA which were “coincidentally” tagmented at the same position by different Tn5 transposomes with different barcodes (Figure 2B). These findings also suggest that the Tn5 tagmentation process is not perfectly random.^{20,23} To reduce these effects in future implementations, one can either increase barcode complexity or reduce input DNA. Using TOM, we can link 4.5% of the fragments, with the longest linkage reaching 1112 bp (Figure S3). This result verified that even using TOM with low complexity barcodes, the combination of barcodes with 9-bp end-duplication at cut-site sequences can be used as UFIs to identify neighbor fragments.

In some applications such as de novo assembly,^{14,29} mapping to high-sequence-similarity genomes³ and RNA sequencing with abundant transcripts,³⁰ high barcode complexity is still required even with the information provided by 9-bp end-duplicate sequences. For these applications, the split-and-pool on-bead DNA synthesis strategy can be employed.³¹ Conventionally, oligonucleotides are synthesized using the 3'-to-5' phosphoramidite chemistry,^{32,33} which produces oligonucleotides with their 3' end attached to the surface. Because the Tn5 transposase ligates the 3' terminus of each transposon strand to target DNA by 3'-OH groups mediated nucleophilic attacks and strand transfer,^{34,35} oligonucleotides with the 3' end attached to the beads rely on their opposite strand to be linked to the target DNA. We hence employed a different immobilization scheme by tethering 3' end of the oligonucleotides to the microbeads to verify whether the orientation of the immobilized oligonucleotides and the additional synthesis step for the opposite strand would affect tagmentation. To generate these functional sequences, we performed high-fidelity 5'-3' elongation initiated by a forward elongation (FE) primer (Figure 3A,B). We measured the reaction efficiency using a FAM labeled FE primer and a TAMRA-labeled probe. Replacement of the TAMRA signal by the FAM signal due to elongation-induced strand displacement indicates the gener-

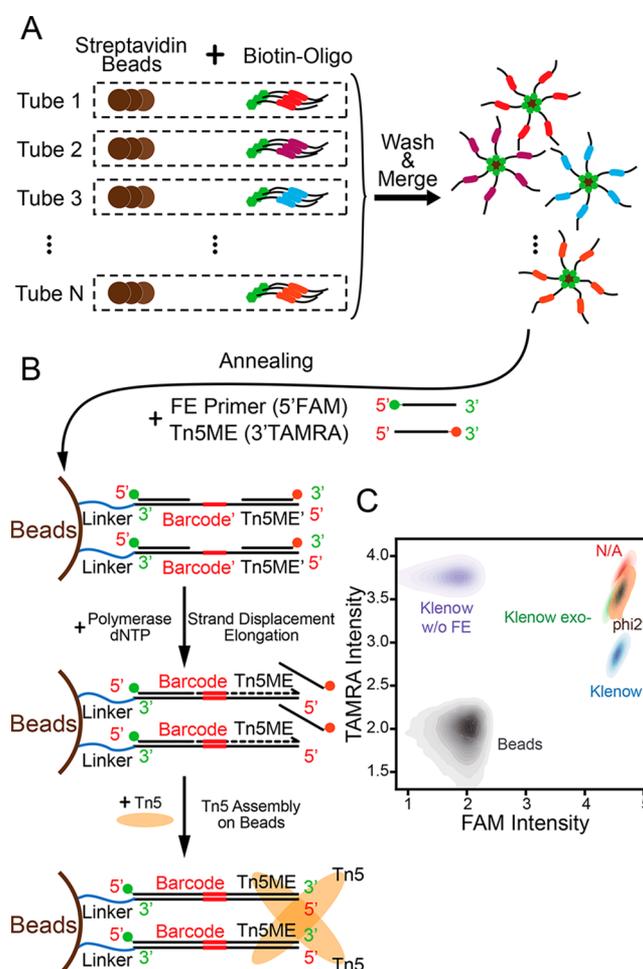


Figure 3. Workflow for 3'-end tethered beads construction and flow cytometry analysis. (A) Workflow for construction of barcoded beads. Each barcoded oligonucleotide was tethered on beads independently. (B) Workflow for Tn5 transposome on beads construction. Two fluorescently labeled primers were annealed to the on-bead oligonucleotides. 5'-3' elongation was performed to generate the forward-strand barcode and Tn5ME sequence, followed by on-bead Tn5 transposome assembly. (C) Kernel density estimate plots of flow cytometry results in log scale. Beads: beads without any primer. N/A: beads with two primers but not elongated by polymerase.

ation of functional double strand barcode and Tn5ME. Thus, the reaction efficiency could be quantitatively assessed by the signals of the two-color fluorescence on microbeads through flow cytometry (Figure 3C and Figure S4). Because the steric effect on the bead surface may have affected the elongation, we compared elongation efficiency among various polymerases with strand displacement property, including Klenow fragment of DNA Polymerase I (Klenow), Klenow fragment exo- and Phi29, and we found Klenow with the highest efficiency. In a control experiment without adding FE primer (Klenow w/o FE), neither elongation nor strand displacement could occur; thus, the microbeads exhibited a high TAMRA signal (Figure 3C and Figure S4). These results verified that elongation occurred on the surface of the beads, and the 5'-3' barcode with Tn5ME had been successfully generated.

We further tested beads of different materials, sizes, and surface density of functional groups to check whether these factors could affect the tagmentation result (Figure 4A). As the result shows, all prevalent beads have comparable results.

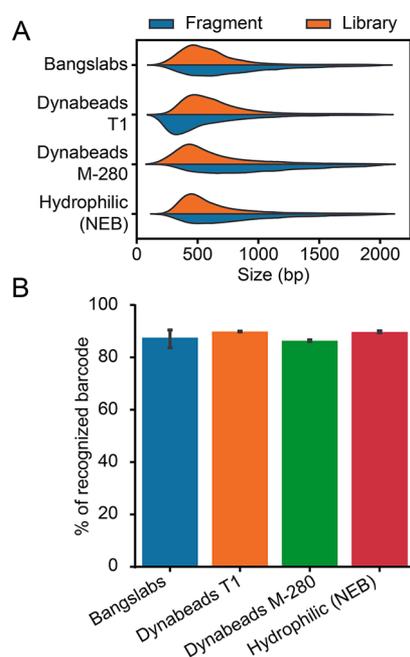


Figure 4. TOM performance using different microbeads. (A) Fragment-size distribution after tagmentation (blue) and library generation (yellow) using different microbeads. (B) Percentage of the fragments that have recognized barcodes in both ends while using 3'-end tethered beads and 5'-3' elongation.

Tagmentation resulted in a broad distribution of fragment size, and after library construction and size selection, the mean size is ~ 450 bp, which is compatible with current NGS platforms (Figure 4A). Because of the high-fidelity elongation, more than 80% of the barcodes at both ends of paired-end reads were

recognized (Figure 4B). Our demonstration showed that oligonucleotides tethered to beads by 3'-end can be used to create high-quality double-strand DNA for barcoding tagmentation. The compatibility of different beads indicated the potential of applying ultrahigh barcode complexity DNA synthesis.

To link the fragments and reconstruct the sequence of original template DNA, a high-reaction-efficiency procedure with low bias and low tolerance for fragment loss is required, which highly depends on barcode design. In our previous design, both ends of the fragments have the same PCR handles, barcodes, and Tn5ME sequences. These long (54 bp) sequences may cause intramolecule hairpin formation during PCR (Figure S5A), and significantly lower the amplification efficiency and introduce bias. Capillary electrophoresis analysis showed that when using beads with mixed barcodes, the size distribution of the fragments was shifted downward compared to those using identical barcoded beads (Figure 5A). This shift suggests an amplification bias toward longer molecules when using identical barcoded beads. This bias might result from the intramolecule hairpin for short fragments. To eliminate this bias, we designed a “wobble” version of a barcode by introducing random nucleotides into the barcode sequence at specific positions to prevent hairpin formation (Figure S5A) with optimized annealing temperature. After this optimization, the fragment size distribution was similar between experiments with mixed and identical barcoded beads. Because the random nucleotides were only located in certain positions in the barcode, we can still accurately identify more than 80% of the barcodes based on their context (Figure S5B). Furthermore, we tested whether different barcodes would introduce sequence-dependent amplification bias by using three different sets of barcodes. We found even distributions of these barcodes, which

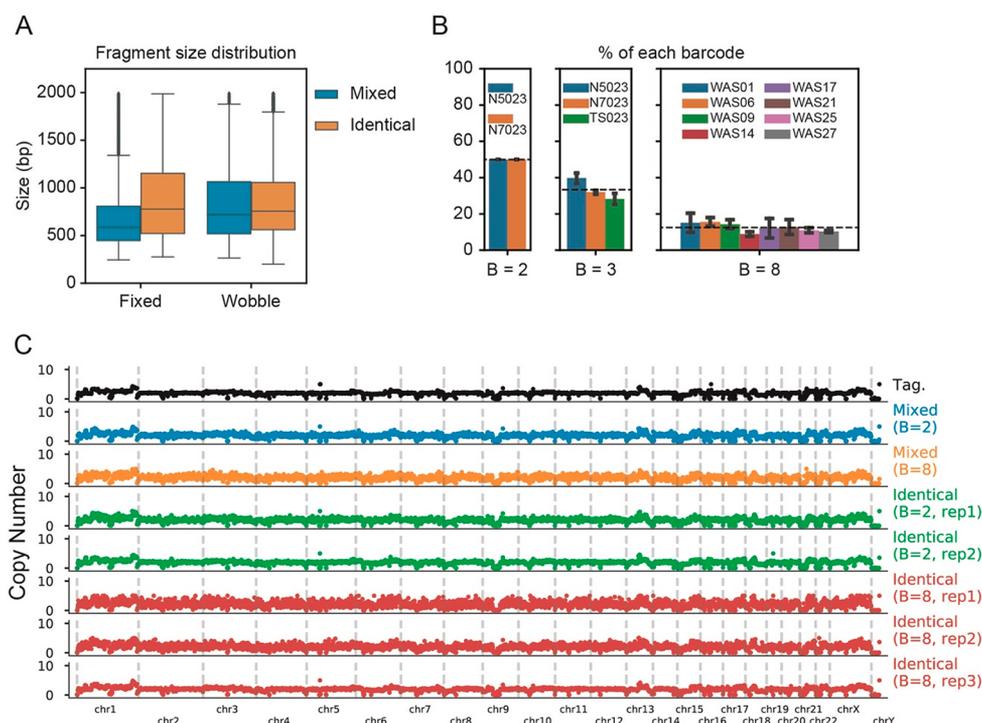


Figure 5. Bias analysis. (A) Boxplot of the fragments size distribution when using two types of beads (Mixed and Identical) and two types of barcodes (Fixed and Wobble). (B) Percentage of each barcode presented in fragments when using different numbers of barcodes (labeled as B). Dashed line represents the expectation. (C) Whole genome copy number profiles using TOM.

indicates no amplification bias was introduced (Figure 5B). We also verified TOM has no sequence dependent bias at whole genome level by performing CNV analysis in a 1M-bin resolution (Figure 5C). When compared with solution-phase Tn5 tagmentation, our method shows highly reproducible and low noise at the whole genome level.

CONCLUSIONS

We developed a method, TOM, to restore long-range information by a sequencing library that was created through tagmentation and barcoding with microbead-immobilized Tn5 transposome. We showed that TOM has acceptable efficiencies for both 5'- and 3'-ends tethered oligonucleotides on microbeads. Tethering oligonucleotides with 5'-ends can construct mid/high complexity barcodes, whereas the 3'-end can be more suitable for ultrahigh-complexity barcodes that required split-and-pool DNA synthesis. We demonstrated that this method could reconstruct template DNA molecule by linking adjacent reads through UFIs, which combine uniquely designed barcodes with the 9-bp duplications at the cut-sites. By using TOM, 4.5% of the sequenced fragments could be linked with their neighbor, and the longest linkage is over 1112 bp.

We are the first to show that the Tn5 transposition generated 9-bp duplications at cut-sites can be used to reduce the barcode complexity requirement for stitching adjacent reads. TOM can be widely used in various sequencing applications, such as haplotype phasing, structure variations detection, alternative splicing analysis, and only requires low complexity of barcodes. Our method could be further improved by increasing the template DNA capture efficiency with the Tn5 transposome on beads and by employing a barcode system with higher complexity. In general, there was a different requirement trade-off between barcode complexity and capture efficiency, which should be taken into consideration for different applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsami.8b01560.

Detailed experimental process, bioinformatics analysis, and Figures S1–S5 (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail for J.W.: jianbinwang@mail.tsinghua.edu.cn.

*E-mail for Y.H.: yanyi@pku.edu.cn.

ORCID

Yanyi Huang: 0000-0002-7297-1266

Author Contributions

H.C., J.W., and Y.H. conceived the project. J.Y. performed the protein purification. H.C. performed the experiments and analyzed the data. H.C., Y.F., Y.H., and J.W. wrote the manuscript. All authors contributed to the discussion and reviewed the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Shuang Geng for discussions and comments at primary stage of the project. We also thank Bo Wu for help with the protein purification. This work was supported by the Ministry of Science and Technology of China (2016YFC0900100 for J.W. and Y.P.), and the National Natural Science Foundation of China (21327808, 21525521 to Y.H. and 21675098 for J.W.).

REFERENCES

- (1) Mardis, E. R. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9* (1), 387–402.
- (2) van Dijk, E. L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. Ten Years of Next-Generation Sequencing Technology. *Trends Genet.* **2014**, *30* (9), 418–426.
- (3) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945.
- (4) Alkan, C.; Sajjadian, S.; Eichler, E. E. Limitations of Next-Generation Genome Sequence Assembly. *Nat. Methods* **2011**, *8* (1), 61–65.
- (5) Peters, B. A.; Kermani, B. G.; Sparks, A. B.; Alferov, O.; Hong, P.; Alexeev, A.; Jiang, Y.; Dahl, F.; Tang, Y. T.; Haas, J.; Robasky, K.; Zaranek, A. W.; Lee, J.-H.; Ball, M. P.; Peterson, J. E.; Perazich, H.; Yeung, G.; Liu, J.; Chen, L.; Kennemer, M. I.; Pothuraju, K.; Konvicka, K.; Tsoumpko-Sitnikov, M.; Pant, K. P.; Ebert, J. C.; Nilsen, G. B.; Baccash, J.; Halpern, A. L.; Church, G. M.; Drmanac, R. Accurate Whole-Genome Sequencing and Haplotyping From 10 to 20 Human Cells. *Nature* **2012**, *487* (7406), 190–195.
- (6) Shalek, A. K.; Satija, R.; Adiconis, X.; Gertner, R. S.; Gaublomme, J. T.; Raychowdhury, R.; Schwartz, S.; Yosef, N.; Malboeuf, C.; Lu, D.; Trombetta, J. J.; Gennert, D.; Gnirke, A.; Goren, A.; Hacohen, N.; Levin, J. Z.; Park, H.; Regev, A. Single-Cell Transcriptomics Reveals Bimodality in Expression and Splicing in Immune Cells. *Nature* **2013**, *498* (7453), 236–240.
- (7) Loman, N. J.; Quick, J.; Simpson, J. T. A Complete Bacterial Genome Assembled De Novo Using Only Nanopore Sequencing Data. *Nat. Methods* **2015**, *12* (8), 733–735.
- (8) Chaisson, M. J. P.; Huddleston, J.; Dennis, M. Y.; Sudmant, P. H.; Malig, M.; Hormozdiari, F.; Antonacci, F.; Surti, U.; Sandstrom, R.; Boitano, M.; Landolin, J. M.; Stamatoyannopoulos, J. A.; Hunkapiller, M. W.; Korlach, J.; Eichler, E. E. Resolving the Complexity of the Human Genome Using Single-Molecule Sequencing. *Nature* **2015**, *517* (7536), 608–611.
- (9) Koren, S.; Schatz, M. C.; Walenz, B. P.; Martin, J.; Howard, J. T.; Ganapathy, G.; Wang, Z.; Rasko, D. A.; McCombie, W. R.; Jarvis, E. D.; Koren, S.; Phillippy, A. M. Hybrid Error Correction and De Novo Assembly of Single-Molecule Sequencing Reads. *Nat. Biotechnol.* **2012**, *30* (7), 693–700.
- (10) Mostovoy, Y.; Levy-Sakin, M.; Lam, J.; Lam, E. T.; Hastie, A. R.; Marks, P.; Lee, J.; Chu, C.; Lin, C.; Dzakula, Z.; Cao, H.; Schlegelbusch, S. A.; Giorda, K.; Schnall-Levin, M.; Wall, J. D.; Kwok, P.-Y. A Hybrid Approach for De Novo Human Genome Sequence Assembly and Phasing. *Nat. Methods* **2016**, *13*, 587–590.
- (11) Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J.; Barnes, C. L.; Bignell, H. R.; Boutell, J. M.; Bryant, J.; Carter, R. J.; Keira Cheetham, R.; Cox, A. J.; Ellis, D. J.; Flatbush, M. R.; Gormley, N. A.; Humphray, S. J.; Irving, L. J.; Karbelashvili, M. S.; Kirk, S. M.; Li, H.; Liu, X.; Maisinger, K. S.; Murray, L. J.; Obradovic, B.; Ost, T.; Parkinson, M. L.; Pratt, M. R.; Rasolonjatovo, I. M. J.; Reed, M. T.; Rigatti, R.; Rodighiero, C.; Ross, M. T.; Sabot, A.; Sankar, S. V.; Scally, A.; Schroth, G. P.; Smith, M. E.; Smith, V. P.; Spiridou, A.; Torrance, P. E.; Tzovev, S. S.; Vermaas, E. H.; Walter, K.; Wu, X.; Zhang, L.; Alam, M. D.; Anastasi, C.; Aniebo, I. C.; Bailey, D. M. D.; Bancarz, I. R.; Banerjee, S.; Barbour, S. G.; Baybayan, P. A.; Benoit, V. A.; Benson, K. F.; Bevis, C.; Black, P. J.; Boodhun, A.; Brennan, J. S.; Bridgham, J. A.; Brown, R. C.; Brown, A. A.; Buermann, D. H.; Bundu, A. A.; Burrows,

- J. C.; Carter, N. P.; Castillo, N.; Chiara E. Catenazzi, M.; Chang, S.; Neil Cooley, R.; Crake, N. R.; Dada, O. O.; Diakoumakos, K. D.; Dominguez-Fernandez, B.; Earnshaw, D. J.; Egbujor, U. C.; Elmore, D. W.; Etchin, S. S.; Ewan, M. R.; Fedurco, M.; Fraser, L. J.; Fuentes Fajardo, K. V.; Scott Furey, W.; George, D.; Gietzen, K. J.; Goddard, C. P.; Golda, G. S.; Granieri, P. A.; Green, D. E.; Gustafson, D. L.; Hansen, N. F.; Harnish, K.; Haudenschild, C. D.; Heyer, N. I.; Hims, M. M.; Ho, J. T.; Horgan, A. M.; Hoschler, K.; Hurwitz, S.; Ivanov, D. V.; Johnson, M. Q.; James, T.; Huw Jones, T. A.; Kang, G.-D.; Kerelska, T. H.; Kersey, A. D.; Khrebtukova, I.; Kindwall, A. P.; Kingsbury, Z.; Kokko-Gonzales, P. I.; Kumar, A.; Laurent, M. A.; Lawley, C. T.; Lee, S. E.; Lee, X.; Liao, A. K.; Loch, J. A.; Lok, M.; Luo, S.; Mammen, R. M.; Martin, J. W.; McCauley, P. G.; McNitt, P.; Mehta, P.; Moon, K. W.; Mullens, J. W.; Newington, T.; Ning, Z.; Ling Ng, B.; Novo, S. M.; O'Neill, M. J.; Osborne, M. A.; Osnowski, A.; Ostadan, O.; Paraschos, L. L.; Pickering, L.; Pike, A. C.; Pike, A. C.; Chris Pinkard, D.; Pliskin, D. P.; Podhasky, J.; Quijano, V. J.; Raczy, C.; Rae, V. H.; Rawlings, S. R.; Chiva Rodriguez, A.; Roe, P. M.; Rogers, J.; Rogert Bacigalupo, M. C.; Romanov, N.; Romieu, A.; Roth, R. K.; Rourke, N. J.; Ruediger, S. T.; Rusman, E.; Sanches-Kuiper, R. M.; Schenker, M. R.; Seoane, J. M.; Shaw, R. J.; Shiver, M. K.; Short, S. W.; Sizto, N. L.; Sluis, J. P.; Smith, M. A.; Ernest Sohna Sohna, J.; Spence, E. J.; Stevens, K.; Sutton, N.; Szajkowski, L.; Tregidgo, C. L.; Turcatti, G.; vandeVondele, S.; Verhovskiy, Y.; Virk, S. M.; Wakelin, S.; Walcott, G. C.; Wang, J.; Worsley, G. J.; Yan, J.; Yau, L.; Zuerlein, M.; Rogers, J.; Mullikin, J. C.; Hurler, M. E.; McCooke, N. J.; West, J. S.; Oaks, F. L.; Lundberg, P. L.; Klenerman, D.; Durbin, R.; Smith, A. J. Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry. *Nature* **2008**, *456* (7218), 53–59.
- (12) Venter, J. C.; et al. The Sequence of the Human Genome. *Science* **2001**, *291* (5507), 1304–1351.
- (13) Lieberman-Aiden, E.; van Berkum, N. L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B. R.; Sabo, P. J.; Dorschner, M. O.; Sandstrom, R.; Bernstein, B.; Bender, M. A.; Groudine, M.; Gnirke, A.; Stamatoyannopoulos, J.; Mirny, L. A.; Lander, E. S.; Dekker, J. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326* (5950), 289–293.
- (14) Burton, J. N.; Adey, A.; Patwardhan, R. P.; Qiu, R.; Kitzman, J. O.; Shendure, J. Chromosome-Scale Scaffolding of De Novo Genome Assemblies Based on Chromatin Interactions. *Nat. Biotechnol.* **2013**, *31* (12), 1119–1125.
- (15) Putnam, N. H.; O'Connell, B. L.; Stites, J. C.; Rice, B. J.; Blanchette, M.; Calef, R.; Troll, C. J.; Fields, A.; Hartley, P. D.; Sugnet, C. W.; Haussler, D.; Rokhsar, D. S.; Green, R. E. Chromosome-Scale Shotgun Assembly Using an In Vitro Method for Long-Range Linkage. *Genome Res.* **2016**, *26* (3), 342–350.
- (16) Kuleshov, V.; Xie, D.; Chen, R.; Pushkarev, D.; Ma, Z.; Blauwkamp, T.; Kertesz, M.; Snyder, M. Whole-Genome Haplotyping Using Long Reads and Statistical Methods. *Nat. Biotechnol.* **2014**, *32* (3), 261–266.
- (17) Fan, H. C.; Wang, J.; Potanina, A.; Quake, S. R. Whole-Genome Molecular Haplotyping of Single Cells. *Nat. Biotechnol.* **2011**, *29* (1), 51–57.
- (18) Schwartz, J. J.; Lee, C.; Hiatt, J. B.; Adey, A.; Shendure, J. Capturing Native Long-Range Contiguity by in Situ Library Construction and Optical Sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (46), 18749–18754.
- (19) Zheng, G. X. Y.; Lau, B. T.; Schnall-Levin, M.; Jarosz, M.; Bell, J. M.; Hindson, C. M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D. A.; Merrill, L.; Terry, J. M.; Mudivarti, P. A.; Wyatt, P. W.; Bharadwaj, R.; Makarewicz, A. J.; Li, Y.; Belgrader, P.; Price, A. D.; Lowe, A. J.; Marks, P.; Vurens, G. M.; Hardenbol, P.; Montesclaros, L.; Luo, M.; Greenfield, L.; Wong, A.; Birch, D. E.; Short, S. W.; Bjornson, K. P.; Patel, P.; Hopmans, E. S.; Wood, C.; Kaur, S.; Lockwood, G. K.; Stafford, D.; Delaney, J. P.; Wu, I.; Ordonez, H. S.; Grimes, S. M.; Greer, S.; Lee, J. Y.; Belhocine, K.; Giorda, K. M.; Heaton, W. H.; McDermott, G. P.; Bent, Z. W.; Meschi, F.; Kondov, N. O.; Wilson, R.; Bernate, J. A.; Gauby, S.; Kindwall, A.; Bermejo, C.; Fehr, A. N.; Chan, A.; Saxonov, S.; Ness, K. D.; Hindson, B. J.; Ji, H. P. Haplotyping Germline and Cancer Genomes with High-Throughput Linked-Read Sequencing. *Nat. Biotechnol.* **2016**, *34* (3), 303–311.
- (20) Adey, A.; Morrison, H. G.; Asan, X.; Kitzman, J. O.; Turner, E. H.; Stackhouse, B.; MacKenzie, A. P.; Caruccio, N. C.; Zhang, X.; Shendure, J. Rapid, Low-Input, Low-Bias Construction of Shotgun Fragment Libraries by High-Density In Vitro Transposition. *Genome Biol.* **2010**, *11* (12), R119.
- (21) Picelli, S.; Björklund, Å. K.; Reinius, B.; Sagasser, S.; Winberg, G.; Sandberg, R. Tn5 Transposase and Tagmentation Procedures for Massively Scaled Sequencing Projects. *Genome Res.* **2014**, *24* (12), 2033–2040.
- (22) Goryshin, I. Y.; Reznikoff, W. S. Tn5 In Vitro Transposition. *J. Biol. Chem.* **1998**, *273* (13), 7367–7374.
- (23) Shevchenko, Y.; Bouffard, G. G.; Butterfield, Y.; Blakesley, R. W.; Hartley, J. L.; Young, A. C.; Marra, M. A.; Jones, S.; Touchman, J. W.; Green, E. D. Systematic Sequencing of cDNA Clones Using the Transposon Tn5. *Nucleic Acids Res.* **2002**, *30* (11), 2469–2477.
- (24) Buenrostro, J. D.; Giresi, P. G.; Zaba, L. C.; Chang, H. Y.; Greenleaf, W. J. Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. *Nat. Methods* **2013**, *10* (12), 1213–1218.
- (25) Amini, S.; Pushkarev, D.; Christiansen, L.; Kostem, E.; Royce, T.; Turk, C.; Pignatelli, N.; Adey, A.; Kitzman, J. O.; Vijayan, K.; Ronaghi, M.; Shendure, J.; Gunderson, K. L.; Steemers, F. J. Haplotype-Resolved Whole-Genome Sequencing by Contiguity-Preserving Transposition and Combinatorial Indexing. *Nat. Genet.* **2014**, *46* (12), 1343–1349.
- (26) Reznikoff, W. S.; Bhasin, A.; Davies, D. R.; Goryshin, I. Y.; Mahnke, L. A.; Naumann, T.; Rayment, I.; Steiniger-White, M.; Twining, S. S. Tn5: a Molecular Window on Transposition. *Biochem. Biophys. Res. Commun.* **1999**, *266* (3), 729–734.
- (27) Reznikoff, W. S. Transposon Tn 5. *Annu. Rev. Genet.* **2008**, *42* (1), 269–286.
- (28) Zhang, F.; Christiansen, L.; Thomas, J.; Pokholok, D.; Jackson, R.; Morrell, N.; Zhao, Y.; Wiley, M.; Welch, E.; Jaeger, E.; Granat, A.; Norberg, S. J.; Halpern, A.; C. Rogert, M.; Ronaghi, M.; Shendure, J.; Gormley, N.; Gunderson, K. L.; Steemers, F. J. Haplotype Phasing of Whole Human Genomes Using Bead-Based Barcode Partitioning in a Single Tube. *Nat. Biotechnol.* **2017**, *35* (9), 852–857.
- (29) Nagarajan, N.; Pop, M. Sequence Assembly Demystified. *Nat. Rev. Genet.* **2013**, *14* (3), 157–167.
- (30) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5* (7), 621–628.
- (31) Macosko, E. Z.; Basu, A.; Satija, R.; Nemes, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A. R.; Kamitaki, N.; Martersteck, E. M.; Trombetta, J. J.; Weitz, D. A.; Sanes, J. R.; Shalek, A. K.; Regev, A.; McCarroll, S. A. Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **2015**, *161* (5), 1202–1214.
- (32) Caruthers, M. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* **1985**, *230* (4723), 281–285.
- (33) Fodor, S.; Read, J.; Pirrung, M.; Stryer, L.; Lu, A.; SOLAS, D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* **1991**, *251* (4995), 767–773.
- (34) Bhasin, A.; Goryshin, I. Y.; Steiniger-White, M.; York, D.; Reznikoff, W. S. Characterization of a Tn5 Pre-Cleavage Synaptic Complex. *J. Mol. Biol.* **2000**, *302* (1), 49–63.
- (35) Bhasin, A.; Goryshin, I. Y.; Reznikoff, W. S. Hairpin Formation in Tn5 Transposition. *J. Biol. Chem.* **1999**, *274* (52), 37021–37029.

Supporting Information

Tagmentation on microbeads: restore long-range DNA sequence information using next generation sequencing with library prepared by surface- immobilized transposomes

He Chen,[†] Jiacheng Yao,[‡] Yusi Fu,[†] Yuhong Pang,[†] Jianbin Wang,^{,‡} and Yanyi Huang^{*,†,§}*

[†]Beijing Advanced Innovation Center for Genomics (ICG), Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[‡]School of Life Sciences, and Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China

[§]College of Engineering, and College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

* E-mail: jianbinwang@mail.tsinghua.edu.cn (J.W.), or yanyi@pku.edu.cn (Y.H.)

Materials and Methods

DNA oligonucleotides and sequences

Barcodes:

WAS01: 5'-WASATCAACGNS-3'

WAS06: 5'-WASGCCAAATNS-3'

WAS09: 5'-WASGATACAGNS-3'

WAS14: 5'-WASAGTATCCNS-3'

WAS17: 5'-WASGTAAGAGNS-3'

WAS21: 5'-WASGTTATCGNS-3'

WAS25: 5'-WASACTAGATNS-3'

WAS27: 5'-WASATTACCTNS-3'

N5023: 5'-CTCTCTATTATC-3'

N7023: 5'-CGTACTAGAGGC-3'

TS023: 5'-CGATGTCAGATC-3'

For 5'-end tethered beads: underlined represents barcode bases

On-beads: 5'-BiotinTEG-Spacer18-

TTTGTGAUGCGATGAACTCAGAGTGCTTNNNNNNNNNNNNNAGATGTGTATAAGAGACAG-3'

Tn5MErev: 5'-Phos-CTGTCTCTTATAACACATCT-NH₂C₆-3'

For 3'-end tethered beads: underlined represents barcode bases

On-beads: 5'-Phos-

CTGTCTCTTATAACACATCTNNNNNNNNNNNNNAAGCACTCTGAGTTCATCGCATCACA
AA-Spacer18-BiotinTEG-3'

Forward elongation (FE) primer: 5'-FAM-TTTGUGATGCGATGAACTCAGAG-3'

Tn5MET-TAMRA: 5'-AGATGTGTATAAAGAGACAGT-TAMRA-3'

PCR primer:

BtgZI_PCR: 5'-GATGCGATGAACTCAGAGTGCTT-3'

Tn5 protein purification

Tn5 transposase was purified as previously described²¹, with the following modification: 12-hour IPTG induction instead of 4-hour induction; protein was bound to the chitin resin by overnight rotating in a centrifuge tube.

Tn5 transposome *In vitro* assembly on barcoded microbeads

Magnetic streptavidin beads (Hydrophilic Streptavidin Magnetic Beads (NEB), Dynabeads MyOne Streptavidin T1 (Invitrogen), Dynabeads M-280 Streptavidin (Invitrogen), Streptavidin Coated Magnetic Classical Particles in 1 μm or 3 μm (Bangs Laboratories, CM01N)) were washed and diluted according to their binding ability by WB buffer (20 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, 0.0005% Triton-X100). 2 μL of 10 μM annealed oligonucleotides with biotin were added to 100 μL beads suspension and incubated at room temperature for 5 min with gentle rotation. Beads were separated by a magnetic separator and washed twice by 100 μL WB buffer to remove unbound oligonucleotides. Functional Tn5ME sequence was already formed on the 5'-end tethered beads after annealing. The 3'-end tethered beads were washed twice by TE buffer, then resuspended in 50 μL elongation buffer containing 1 μL polymerase (Klenow, Klenow exo-, Phi29, purchased from NEB), 1 mM dNTP and 1X reaction buffer (NEBuffer 2 for Klenow and Klenow exo-, Phi29 reaction buffer for Phi29), and incubated at reaction temperature (37°C for

Klenow and Klenow exo-, 30°C for Phi29) for 30 min with gentle rotation. After elongation, beads were separated by a magnetic separator and washed twice by 50 µL WB buffer. Both 5'- and 3'-end tethered beads with functional Tn5ME sequence were resuspended by 100 µL WB buffer and stored at 4°C. Before use, beads were separated by magnet and wash twice by 20 µL TE buffer. Upon usage, 5 µL purified Tn5 transposase was added to it with a few pipetting and incubated at room temperature for 30 min with gentle rotation.

Tn5 transposome on microbeads tagmentation and gap-filling

For each 5 µL on-beads Tn5 transposome, 10 µL H₂O, 1 µL HEK293 cell gDNA (5 ng/µL) and 4 µL 5X TAPS buffer (50 mM TAPS-NaOH, 25 mM MgCl₂, 40% PEG-8000) were added. Tagmentation was performed at 55°C for 1 hour, which has better performance than 10 min and similar performance as 3 hours (data not shown), with gentle rotation. After tagmentation, the target DNA were linked with barcode beads by either covalent bond (5'-end tethered beads) or hydrogen bond (3'-end tethered beads). The reaction was stopped by magnet separation. The beads were wash twice by 20 µL Strip buffer (20 mM EDTA, 5 mM Tris-HCl) and incubated in 20 µL Strip buffer at 50°C for 30 min to strip The Tn5 from target DNA. Then the beads were separated by magnet and washed twice by 20 µL TE buffer.

For 5'-end tethered beads, gap-filling buffer (15 µL H₂O, 2 µL NEBuffer 4, 2 µL of 10 mM dNTP, 1 µL Klenow exo-, 1 µL USER Enzyme (NEB)) were added to resuspended the beads and incubated at 37°C for 30 min. After gap-filling, beads were separated by magnet and supernatant was collected. To prevent the free primer messed up the barcode during PCR, digestion was performed by adding 1 µL Mung Bean Nuclease (NEB) and incubated at 30°C for 10 min, followed by adding 1 µL 5X TS buffer (Vazyme) to deactivate Mung Bean Nuclease. For 3'-end

tethered beads, gap-fill were performed by melting and Q5 elongation in 10 μ L H₂O, 10 μ L Q5 High-Fidelity 2X Master Mix (NEB) and incubated at 72°C for 30 min.

PCR amplification and purification

For 5'-end tethered beads, each tube was added 5 μ L of 10 μ M Primer and 25 μ L Q5 High-Fidelity 2X Master Mix. For 3'-end tethered beads, each tube was added 10 μ L H₂O, 5 μ L of 10 μ M Primer and 15 μ L Q5 High-Fidelity 2X Master Mix. Then PCR amplified with the following program: 98°C for 3 min, 18 cycles of 98°C for 25 s, 60°C for 30 s and 72°C for 90 s, followed by 72°C for 3 min and hold at 4°C. After PCR, DNA was purified by 1:1 ratio of VAHTS DNA Clean beads (Vazyme), quantified by Qubit dsDNA HS Assay Kit (Invitrogen), and analyzed by Fragment Analyzer (Advanced Analytical Technologies).

Preparation of NGS compatible library and sequencing

A Type II restriction enzyme was used to remove the PCR handle in the both ends of fragments in a 20 μ L reaction contain 10~500 ng DNA, 1X NEBuffer 4, and 1 μ L BtgZI (NEB). After that, standard library preparation was performed according to the standard procedure of NEBNext Ultra II DNA Library Prep Kit (NEB) with 200~600 bp size selection. The libraries were pooled together, qualified and sequenced on Illumina MiniSeq or HiSeq platform.

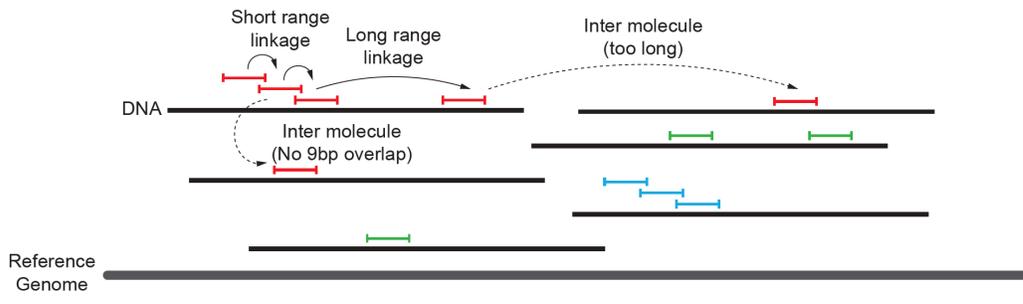
Sequencing data analysis

Sequencing adapters, un-cut PCR handle, and Tn5ME sequence on both ends of paired reads were recognized and removed by Cutadapt (Version 1.12),³⁶ the 12 bp barcodes were recorded. Then reads were aligned to human genome reference hg38 by Bowtie2 (Version 2.2.9) with -N 1

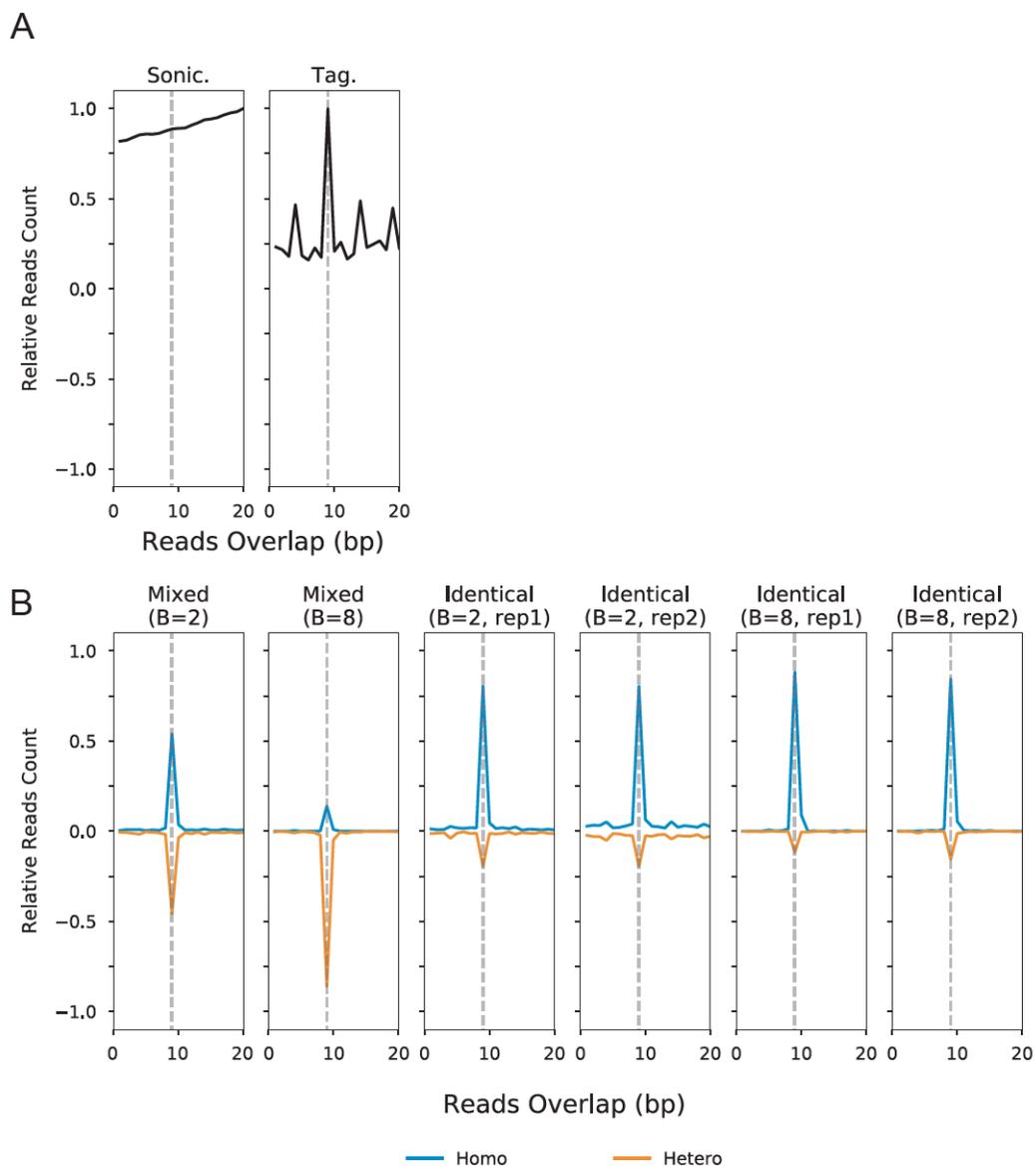
-L 20 -reorder, and stored as a BAM file.³⁷ Then the aligned reads were paired and labeled by barcode at both ends by a custom Perl script, and transformed into a BED file, overlapped fragments were extracted by Bedtools (Version 2.26.0).³⁸ To perform CNV analysis, reads in BAM file were filtered by mapping quality score >20 and duplicates were removed by Samtools (Version 1.3.1),³⁹ then copy number was called using CNVkit package.⁴⁰ Results were plotted by a custom Python script using Pandas and Seaborn package.

FACS analysis

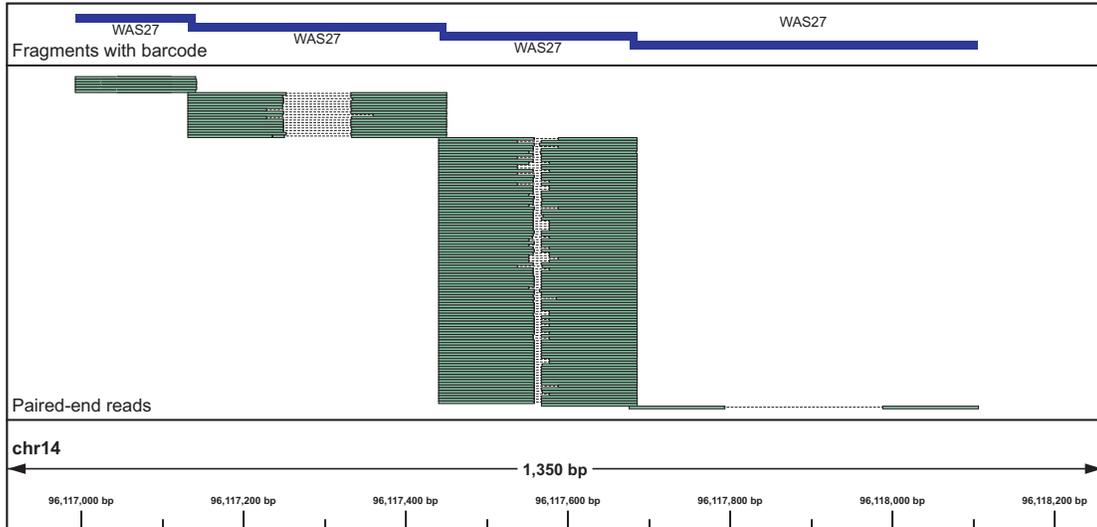
Beads were separated by magnet and washed by WB buffer and resuspended in 0.1X WB buffer. BD LSRFortessa cell analyzers were used to perform fluorescence-activated cell sorting (FACS) analysis. FAM was measured by 488 nm laser with 530/30 filter, TAMRA was measured by 561 nm laser with 585/20 filter. For each sample, more than 10⁴ events were recorded.



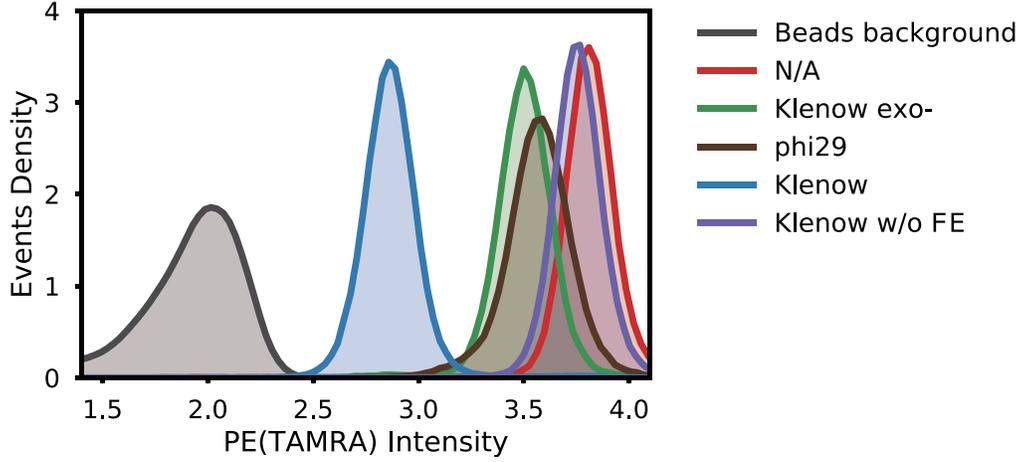
Supplementary Figure S1. Illustration of the long-range linkage using TOM.



Supplementary Figure S2. Relative counts for overlapped fragment pairs. (A) The 9-bp overlapped fragments was enriched in TOM library, but not in sonication fragmented library. (B) Comparison of using mixed beads and identical beads.

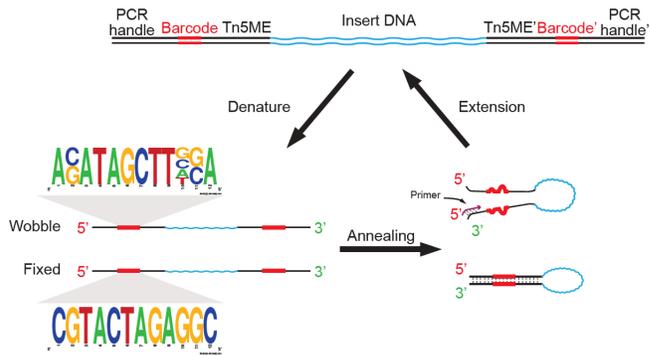


Supplementary Figure S3. A demonstration of the longest linkage.

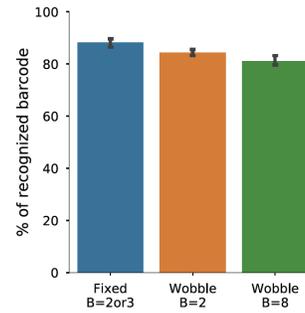


Supplementary Figure S4. Density plots of flow cytometry data in log scale.

A



B



Supplementary Figure S5. Wobble barcode design and recognition. (A) Illustration of the fragment, barcode and hairpin formation. (B) Percentage of the fragments that have recognized barcodes in both ends while using wobble barcodes.

REFERENCES

- (36) Martin, M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet J.* 2011, *17* (1), 10.
- (37) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 2012, *9* (4), 357–359.
- (38) Quinlan, A. R.; Hall, I. M. BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 2010, *26* (6), 841–842.
- (39) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 2009, *25* (16), 2078–2079.
- (40) Talevich, E.; Shain, A. H.; Botton, T.; Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization From Targeted DNA Sequencing. *PLoS Comput. Biol.* 2016, *12* (4), e1004873.